DOCUMENT RESUME

ED 414 289 TM 027 712

AUTHOR King, Jason E.
TITLE Methods of Assessing Replicability in Canonical Correlation Analysis (CCA).
PUB DATE 1997-01-25
NOTE 21p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (20th, Austin, TX, January 25, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Correlation; *Effect Size; *Multivariate Analysis; *Robustness (Statistics); *Sampling
IDENTIFIERS *Bootstrap Methods; *Research Replication; Weighting (Statistical); Wherry Formula

ABSTRACT

Theoretical hypotheses generated from data analysis of a single sample should not be advanced until the replicability issue is treated. At least one of three questions usually arises when evaluating the invariance of results obtained from a canonical correlation analysis (CCA): (1) "Will an effect occur in subsequent studies?"; (2) "Will the size of the effect replicate?"; and (3) "Will variables induce similar weightings on replicated variables?" When "external" replication is not feasible, "internal" invariance estimates using only data in hand can yield information about the preceding questions. Approaches to assessing replicability that are discussed are Wherry's correction formula, rotation of function and structure matrices, crossvalidation, and bootstrapping techniques for CCA. Even the bootstrap method is afflicted with the problem common to all the methods of replicability discussed in this paper: results are based solely on data obtained from a single sample. Since this technique creates multiple configurations of the data, it is more robust to the sample-specificity limitation, but it is always advisable to compute multiple invariance estimates. These approaches can aid in determining the replicability of research results, but they should only be used when external resampling is not feasible or practical. (Contains 21 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running head: ASSESSING REPLICABILITY IN CCA

ED 414 289

Methods of Assessing Replicability in Canonical Correlation Analysis (CCA)

Jason E. King

Texas A&M University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Jason King

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM027712

Paper presented at the 20th annual meeting of the Southwest Educational Research Association, Austin, TX, January 25, 1997.

BEST COPY AVAILABLE

Abstract

Theoretical hypotheses generated from data analysis of a single sample should not be advanced until the replicability issue is treated (Thompson, 1996). At least one of three questions usually arises when evaluating the invariance of results obtained from a canonical correlation analysis (CCA): (a) Will an effect occur in subsequent studies? (b) Will the size of the effect(s) replicate? (c) Will variables induce similar weightings on replicated variates? When "external" replication is not feasible, "internal" invariance estimates using only data in hand can yield information relevant to the above questions. The present paper explains Wherry's correction formula, rotation of function and structure matrices, crossvalidation and bootstrapping techniques for CCA.

Methods of Assessing Replicability in Canonical Correlation Analysis (CCA)

As is well documented, basing theories on results obtained from a single sample often effects inaccurate hypotheses in explaining parametric relationships (Thompson, 1990). To avoid this error, methodologists continue to counsel estimating the replicability of research results (e.g., Thompson, 1995). No doubt one reason for the varied conclusions often reached by researchers studying similar constructs is the sample-specificity of the statistics computed. Each sample collected from a population of interest will yield at least slightly different results from any other independent sample. Thus, two individuals can draw samples from identical populations and yet infer diverse theories based on their data.

Some researchers do not worry about testing result replicability, because they incorrectly believe that statistical significance tests evaluate whether similar results occur in the population, or are replicable. Of course, such fallacies have been thoroughly disproven (Cohen, 1994; Thompson, 1996). Thus, APA has constituted a Task Force on Statistical Inference which will consider, among other recommendations, whether significance tests should be banned from APA journals (Shea, 1996). Methods other than statistical significance tests must be used to evaluate likely result replicability.

Ideally, researchers draw multiple samples from the population and generate theory based on the plurality of results obtained. In lieu of repeated sampling, however, one can appraise "internal" replicability using data from the sample in hand. These types of replicability assessments are useful for many least squares procedures such as linear regression (e.g., Fan & Jacoby, 1995; Neter, Kutner, Nachtsheim, & Washerman, 1996; Thompson, 1994), discriminant analysis (e.g., Dalgleish & Chant, 1995; Thompson, 1993), and factor analysis (e.g., Gorsuch, 1983; Thompson, 1988). This paper will present several methods of estimating the replicability of research results obtained from canonical correlation analysis (CCA).

Basic Steps in Canonical Correlation Analysis

CCA is a general multivariate procedure used to examine the relationships between two sets of variables. Thus, one might be studying results from a survey assessing both social issue opinions (e.g., for/against the death penalty, for/against abortion) and views on educational reforms (e.g., for/against integrated classrooms, for/against the use of new teaching methods). The uninformed researcher might factor analyze all of the variables at once. If one wished to preserve the two distinct groupings of variables, however, CCA would be the appropriate procedure. The specific techniques used to derive the coefficients produced in canonical analysis are complex and are usually performed by computer; however, a brief overview of the

basic workings of CCA will be helpful in understanding the dynamics operating in replicability analyses.

Initially, a correlation matrix (matrix R) containing all variables in the study is partitioned into four quadrants (two "intervariable" and two "intravariable" quadrants). Using matrix algebra, these submatrices are manipulated resulting in a "quadruple-product matrix" (matrix A), while keeping the sets of variables intact (see Thompson, 1984, for details). Eigenvalues are then obtained by performing a principal components analysis on the quadruple-product matrix. The number of resulting eigenvalues will always equal the number of variables in the smallest variable set (Thompson, 1991).

Since all least squares analyses are related (Cohen, 1968; Knapp, 1978), the workings of CCA can be conceptualized more intuitively within a regression framework. An analog to the regression equation is the canonical function equation. Within each set, linear combinations of the variables are formed by weighting (parallel to assigning regression b weights) resulting in synthetic or latent variables called canonical variates. The weights are denoted canonical function coefficients.

The variates are then paired with each other across variable sets. Specifically, the canonical variates "are uncorrelated within and between sets except that they correlate pairwise in decreasing order of magnitude" (Bentler & Huba, 1982, p. 24). A pairwise correlation of variates from each set is a canonical

correlation coefficient (R_c). In fact, the R_c is parallel to the Multiple R in linear regression.

Scores on the latent variables or functions are designated variate scores in CCA (cf., regression yhat scores). As in regression, coefficients that measure the relationship between observed and latent variables are called structure coefficients. Finally, communality coefficients indicate the extent to which each variable contributes to the functions. They are computed by summing the squared structure coefficients of a variable across all functions. Other coefficients are also calculated in CCA but a knowledge of their functioning is not needed to understand the invariance techniques discussed in the present paper.

Estimates of Replicability

Invariance estimates are quantitative methods of determining the replicability of a study's results. If the statistics obtained from an initial analysis approximate in repeated analyses, one is more likely to obtain similar statistics in future samples drawn from the same population. One might argue that invariance estimates are not needed if the initial sample is selected randomly, but this argument ignores the sample-specific nature of obtained statistics. Even randomly selected samples can produce spurious results owing to sampling error. While analyses utilizing the data in hand are helpful in determining if the results will likely replicate, it should be remembered that invariance estimates are just that--estimates, not precise

predictions. Only with external resampling can one confirm that results are truly replicable.

At least one of three questions usually arises when evaluating the invariance of results obtained from a CCA: (a) Will an effect occur in subsequent studies? (b) Will the size of the effect(s) replicate? (c) Will variables induce similar weightings on replicated variates? Each of these issues can be appraised by applying one or more of the techniques presented below. Generally, it is preferable to restrict inferences of replicability to evaluating the presence or absence of an effect (i.e., question [a] above). As more specific data features are tested for invariance (e.g., the values of function coefficients obtained for each variable per function), the probability of detecting similar results in future samples decreases. Nevertheless, the more specific invariance estimates can be helpful in building theory if used in a responsible manner.

Wherry's Correction Formula

Probably the simplest replicability estimate to compute is Wherry's correction formula.

$$\underline{R}_{.}^{2} - ((1-\underline{R}_{.}^{2}) * (\underline{V}/(\underline{N}-\underline{V}-1)))$$

where $\underline{V}$=total number of variables

While it may not usually be adjudged an invariance estimate, the calculation does provide information as to the replicability of results. It is beneficial in determining the replicability of the presence or amount of effect size quickly (i.e., questions [a] or

[b]). Since samples tend to slightly overestimate population parameters, this correction was developed to attenuate the distortion. The data are not resampled or manipulated in any way, however, so this is not a powerful method of assessing replicability. Only the canonical correlation coefficient (R), sample size (N), and number of variables (V) are employed in the formula.

Rotation

The technique of varimax rotation is associated primarily with principal components analysis (PCA) and factor analysis (FA) in the minds of many. However, as mentioned above, PCA is one of the steps in CCA as well. Some have suggested that rotation can also contribute to assessing the replicability of the data. Kuylen and Verhallen state, "A solution is said to be invariant when the same groups of variables are found repeatedly on variates whenever at least some important variables, which reflect the underlying construct, are used in the repeated analysis" (cited in Thompson, 1984, p. 35). Thus, obtaining functions that are composed of the same variables over multiple rotations would imply that the data results are replicable. Rotation can potentially yield information pertinent to all three of the questions commonly asked in a replication analysis.

It might seem that rotation consists of manipulating the data until it takes the form that we wish it to have. But this is not true since the communality coefficient of each variable

remains fixed during rotation. Furthermore, the sums of squared canonical correlation coefficients across functions will remain constant for pre- and post-rotation results. In essence, a redistribution of the variance across the functions is the only change that takes place during the procedure (though all coefficients should be recomputed after rotation).

Yet there are some problems to consider before rotating the canonical functions. First, what matrix should be rotated? The function coefficient matrix, for example, could be rotated. This would provide information as to the replicability of the raw weights applied to each variable on each function of interest. Since the weights are not standardized and can be greater than one, however, interpretation becomes more difficult. It is generally easier to understand and explain standardized coefficients. Another reason for not rotating the function coefficients is that the sum of squared weights across columns does not result in a meaningful statistic, unlike communality coefficients obtained from adding the squared structure coefficients. Because of these two drawbacks, the structure coefficient matrix should usually be rotated in assessing invariance, but the research questions being asked may supersede this generality (viz., there are times when the function matrix should be rotated).

A second problem associated with this type of rotation (i.e., varimax) is that all the variables are "lumped together,"

thus, destroying the two sets of variables that had been isolated. Simplification of results might occur in one set of variables but not the other (due to all variables being treated equally in the rotation). Further, the variable sets are implicitly weighted by the proportion of variables present in each set. Thus, Bentler and Huba (1982) developed a rotation procedure that allows one to specify the relative influence of each variable set in the analysis. Since the variables are weighted according to the set they belong to, rotation will produce results that are more invariant, that is, the rotation procedure is "forced" to consider the variables in pre-determined groupings in factor space. However, the solution no longer yields maximally correlated results because the categorization process restricts the least squares method of determining maximum correlations. The researcher has constrained the rotation to take into account weighted sets of variables instead of considering each variable independently (Thompson, 1984).

Depending on the research questions being asked, rotation may or may not be useful in assessing the replicability of data results. Bentler and Huba (1984) admonish against rotation if a large general factor appears across each set of variables. Rotation will artificially split the general factor into several dimensions. Further, if rotation is used, both "symmetric" and "asymmetric" rotations using different numbers of dimension should be compared (see Bentler & Huba for details).

Crossvalidation

A simpler and more common means of assessing replicability is through crossvalidation. Like rotation, crossvalidation techniques can provide probabilistic answers to all three questions of interest. Based on the results of a CCA, one selects the functions "important" enough to warrant replicability assessment. The function or structure coefficients can then be cross-validated. This paper will present two such methods: partial- and full-model estimates.

Partial-model estimates. In this technique, one first splits the data into two subgroups. Since there are two variable sets in CCA, a splitting of the data into subgroups will result in a total of four subsets (two subsets per variable group). In single crossvalidation, only one of the subgroups is used for the analysis; in double crossvalidation, statistics are computed for both subgroups.

The data can be divided equally (i.e., a 50-50 split) or in some other fashion (e.g., a 30-70 split), but data-splitting should be based on empirical reasons. For example, in a study with a small sample size, a 30-70 split would increase the likelihood of obtaining spurious results in the smaller subset. It is generally advisable not to use crossvalidation techniques with small samples (Taylor, 1992). Instead, a method such as the bootstrap might be considered which is more robust to modest sample size.

For single crossvalidation, new function coefficients (or structure coefficients) are next calculated for one subgroup on each function of interest by applying the function (or structure) coefficients of subset 2 to the variables in subset 1 and vice-versa. If function coefficients are used, values must first be converted into standardized scores (parallel to beta weights in regression). Squared correlations (R_c^2s) are then calculated for each of the variable sets in the subgroup.

Finally, an invariance coefficient is calculated by subtracting the squared canonical correlation of subset 1 from subset 2 for the first subgroup. In double crossvalidation, another invariance coefficient would be calculated from the two variable sets within the second subgroup. If these coefficients are small, the data are said to be invariant. See Taylor (1992) for a sample SPSS program.

Full-model estimates. In the technique discussed above, one splits the data into two subgroups and computes new statistics for the data. Since the invariance results are based on these abbreviated samples, the method is referred to as a partial-model estimate. But it is also possible to use the entire sample in testing replicability. This crossvalidation technique is a full-model estimate because no information pertaining to the entire sample is lost in the calculation process.

Here, the data are first randomly split into two subgroups. After computing a canonical solution for each subgroup, one

solution is arbitrarily rotated (through a varimax rotation) on to the other solution using a "Procrustean" rotation (see Thompson, 1984). Again, either function or structure coefficients can be used in the operation, depending on whether one is interested in the similarity of function equations in producing variate scores or the similarity of the distribution of variable variances across equations. The cosines of the angles among the rotated functions or structures are correlation coefficients. Therefore, the cosines can be treated as invariance estimates: If they are homogeneous and large, the data are potentially replicable.

In an earlier section, it was noted that varimax rotation destroys distinctions between the two variable sets that have been created prior to the analysis. It might seem that the same criticism could be levied against this crossvalidation technique. But in this scenario, one is not interested in interpreting the results of the rotation, only in the correlation coefficients produced through the procedure. There are potential problems to be considered, however. First, there are many possible splits of the sample, all of which may yield slightly different results. Second, more variables will limit the amount of rotation possible. As an increasing number of variables fill the factor space, the positions of the matrix will become more fixed resulting in conservative estimates of replicability (Thompson, 1984). In spite of the drawbacks, crossvalidation is often

useful. But since "invariance testing is a relatively new field and parameters for interpreting the results are just beginning to be established" (Taylor, 1992, p. 12), one should exercise caution in basing complete confidence in the results from a cross-validated study alone.

Bootstrapping

A more powerful method for assessing replicability is the bootstrap technique. Again, this method can be used to answer any of the three questions of interest in a replication study. The bootstrap is an empirical estimation procedure that resamples the data hundreds, even thousands of times, resulting in a "mega" data set (Thompson, 1991). The mega file is then used as a data bank from which to draw multiple samples. Since this procedure involves creating more samples through "internal" calculations without "external" help (i.e., drawing further samples from the population), it is, in effect, "pulling itself up by its own bootstraps."

The newly created samples have sizes identical to the original sample. Each subject in the original data set can be resampled many times or never again because the resampling process is random; thus, each resample results in a unique configuration of the data. This makes for a very powerful procedure.

A CCA is next computed for each resample, followed by an averaging of the results. The average of the squared canonical

correlation coefficients (R_c^2s) can be considered as well as the other coefficients generated from the analysis (i.e., structure and function coefficients). The standard deviations obtained are empirical estimates of the standard error of the statistics and can also be used in making decisions about the replicability of the coefficients obtained (Thompson, 1995).

One difficulty with creating a bootstrap program for canonical analysis is that the functions must remain identical across rotations. For example, consider two variable sets, each containing two variables loading on two functions. If a resampled data set is rotated in a typical fashion, the variables might subsequently load on different functions, and, therefore, the two solutions would not be directly related. Only when functions remain constant across resampling can one legitimately compare results from the multiple samples.

To keep the functions constant, a common factor space is needed to serve as a target solution upon which the functions can be rotated. Either (a) a simple matrix consisting of zeros, ones, and negative ones (for an example applied to discriminant analysis, see Dalgleish & Chant, 1995); (b) a previously discovered matrix isolated in theory; or (c) a structure or weight matrix for the sample data in hand can be used as the target solution (Thompson, 1995). The CANSTRAP (Canonical Bootstrap) computer program written by Thompson (1995) uses the sample data as a target matrix. Willemain (1994) has demonstrated

BEST COPY AVAILABLE

how bootstrap techniques can be employed through the use of spreadsheets.

Conclusion

Even the bootstrap is afflicted with the problem common to all methods of replicability assessment presented in this paper, that is, results are based solely on the data obtained from a single sample. Since this technique creates multiple configurations of the data, it is more robust to the sample-specificity limitation, but it is always advisable to compute multiple invariance estimates. Those presented here can potentially aid in determining the replicability of research results; nevertheless, they should only be employed when external resampling is not feasible or practical.

References

Bentler, P. M., & Huba, G. J. (1982). Symmetric and assymetric rotations in canonical correlational analysis: New methods with drug variable examples. In N. Hirschberg & L. G. Humphreys' (Eds.), Multivariate applications in the social sciences (pp. 21-46). Hillsdale, NJ: Erlbaum.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Dalgleish, L. I., & Chant, D. (1995). A SAS macro for bootstrapping the results of discriminant analysis. Educational and Psychological Measurement, 55, 613-624.

Fan, X., & Jacoby, W. G. (1995). BOOTSREG: An SAS matrix language program for bootstrapping linear regression models. Educational and Psychological Measurement, 55, 764-768.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Liang, K. -H., Krus, D. J., & Webb, J. M. (1995). K-fold crossvalidation in canonical analysis. Multivariate Behavioral Research, 30, 539-545.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied Linear Regression Models (3rd ed.). Chicago, IL: Irwin.

Shea, C. (1996). Psychologists debate accuracy of "significance test." Chronicle of Higher Education, 42(49), A12-A16.

Taylor, D. L. (1992). Cross-validation in canonical analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 31-February 2, 1992. (ERIC Document Reproduction Service No. ED 342 809)

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: Sage.

Thompson, B. (1988). Program FACSTRAP: A program that computes bootstrap estimates of factor structure. Educational and Psychological Measurement, 48, 681-686.

Thompson, B. (1990). Variable importance in multiple regression and canonical correlation. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA, April 16-20, 1990. (ERIC Document Reproduction Service No. ED 317 615)

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24, 80-95.

BEST COPY AVAILABLE

Thompson, B. (1993). DISCRSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62(2), 157-176.

Thompson, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. Educational and Psychological Measurement, 55(1), 84-94.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Wildt, A. R., Lambert, Z. V., & Durand, R. M. (1982). Applying the jackknife statistic in testing and interpreting canonical weights, loadings, and cross-loadings. Journal of Marketing Research, 19, 99-107.

Willemain, T. R. (1994). Bootstrap on a shoestring: Resampling using spreadsheets. The American Statistician, 48(1), 40-42.

Footnotes

[1]Liang, Krus, and Webb (1995) have suggested a K-fold crossvalidation technique. Specifically, the data are split into two sub-samples, cross-validated, normalized and fitted to a running composite. These iterations are then repeated until a specific criterion for termination is attained.

[2]Tukey's jackknife procedure might also be employed to assess replicability (as applied to CCA, see Wildt, Lambert, & Durand, 1982). However, Dalgleish (as cited in Dalgleish & Chant, 1995) found that the jackknife was too liberal in its Type 1 error rates.

BEST COPY AVAILABLE